THE EMBO JOURNAL

Manuscript EMBO-2015-93360

Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking *in vivo*

Erik Holmqvist, Patrick R. Wright, Lei Li, Mr. Thorsten Bischler, Lars Barquist, Richard Reinhardt, Rolf Backofen and Jorg Vögel

Corresponding author: Jörg Vogel, University of Würzburg; Rolf Backofen, Albert-Ludwigs-University of Freiburg

Review timeline:

Submission date:	09 November 2015
Editorial Decision:	03 December 2015
Revision received:	26 January 2016
Editorial Decision:	16 February 2016
Revision received:	25 February 2016
Accepted:	29 February 2016

Editor: Anne Nielsen

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

1st Editorial Decision 03 December 2015

Thank you for submitting your resource manuscript for consideration by the EMBO Journal. It has now been seen by three referees whose comments are shown below.

As you will see from the reports, all referees express interest in the findings reported in your manuscript, although they do raise a number of concerns that you will have to address before they can support publication in The EMBO Journal.

For the revised manuscript I would particularly ask you to focus your efforts on the following points:

-> Please expand the description and analysis of the data sets (ref #1) to more clearly emphasize and utilize the three biological replicates.

-> Please include additional reporter constructs with mutated binding sites to lend further support to the specificity of the regulatory effect (ref#1)

-> Please discuss/clarify the predictive power in target identification as well as the possibility for quantitative information in the present data set as requested by ref #2. This referee furthermore lists a number of minor points on data interpretation and controls that should be addressed.

Given the referees' overall positive recommendations, I would thus like to invite you to submit a revised version of the manuscript, addressing the comments of all three reviewers.

Thank you for the opportunity to consider your work for publication. I look forward to your revision.

REFEREE COMMENTS

Referee #1:

In this manuscript Holmqvist and coworkers determine the RNA binding sites of two major bacterial post-transcriptional regulators, Hfq and CsrA, in Salmonella Typhimurium. Using CLIP-seq they identify RNA preferences and structural constraints of Hfq and CsrA. The binding site maps for Hfq interacting sRNAs and their target mRNAs support a model for Hfq as a mediator of RNA duplex formation, and provide new insight into improving sRNA target prediction.

The manuscript is well written and the results clearly presented. The conclusions are supported by experimental and computational results. The research topic and findings are of broader interest. Before considering this manuscript for publication, some issues should be addressed.

The authors should mention that they performed biological Hfq CLIP replicates. Since multiple replicates were generated the reproducibility and variance of these CLIP replicates should be assessed. This should not only been done at the transcript level, but also for binding sites. Based on the description in the methods sections reads from three replicates were pooled and the peak calling was performed. If this was the case, why not using more stringent criteria to identify peaks that have sequence coverage from at least 2 of the three replicates.

In Figure 2A the authors should also plot the peak density of peaks with containing T>C mutations The authors argue that Hfq RNA binding data could increase the success of sRNA-target mRNA predictions. Using a mglB-gfp reporter they show that Spot42 represses a constitutively transcribed mglB-gfp fusion. To confirm that repression is mediated through an interaction of Spo42 with the mglB site ,identified by CLIP, the authors should mutate the reporter and show a direct effect on mglB-gfp expression.

As for the analysis of the Hfq CLIP-seq data the reproducibility and variance of CsrA CLIP replicates should be assessed. The authors show using a sopD2-gfp reporter that CsrA may directly repress the sopD2-gfp fusion. The reporter assays indicate that CsrA is involved in the regulation. To show that CsrA regulation is mediated by the interactions with the CLIP identified site in sopD2 the authors should mutate (or delete) the GGA motif in the binding site to examine the contribution of this motif.

Lastly, the authors discuss the differences in RNA binding results from their study compared to Hfq CLIP data by Tree et al. at the molecular level. The question that arises is, whether targets are similar at the transcript level in E. coli and Salmonella Typhimurium.

Referee #2:

This manuscript describes the use of CLIP-seq in Salmonella to globally map protein-RNA interactions for two important RNA binding proteins, Hfq and CsrA. The data sets confirm interactions with known mRNA targets of these two proteins, and yields consensus binding site motifs that are consistent with published work. In addition, new putative targets are identified, and the authors do some experimental work to confirm that the binding sites found on some new targets are responsible for physiologically-relevant regulatory effects. The study provides a valuable new experimental model for genome-wide identification of protein-RNA interactions. In addition, the work suggests some general principles relevant to the activities of these two RNA binding proteins, e.g., the frequent location of Hfq binding sites 5' relative to sRNA binding sites on mRNA targets. Moreover, the results suggest that inclusion of Hfq binding site data could be used to improve

computational prediction of sRNA-mRNA interactions. Overall, the study is extremely valuable, well written, and will be of broad interest. I am concerned about the qualitative nature of most of the observations, and think the study would be much improved if the authors could provide some context for or relative quantification of Hfq/CsrA interactions with the many different mRNA targets. My specific comments are below.

• The SPIs are more AT-rich, and this might increase non-specific interaction of Hfq with AU-rich RNAs derived from SPI1. Could this account for increased signal for Hfq across the SPIs?
• The Hfq consensus from peaks in 3' UTRs and sRNAs does resemble half of a GC-rich hairpin followed by a run of Us, as you would find in Rho-independent terminators. How frequently do the peaks containing this consensus contain the upstream half of the hairpin? The density of peaks around the terminator (Fig. 2H) does suggest that most of these motifs represent the real terminator of the sRNA or mRNA, but this was not exactly clear.
• My major concern is the lack of relative quantification of Hfq/CsrA interactions with different targets. The description of how peaks were called (in Materials and Methods) was complex and hard to follow. Some simplified description of what the authors mean by "significant peak" would be very helpful, as would a description of the range of peak size (read abundance?), normalized to total mRNA abundance, perhaps?
• Please comment on the sensitivity of detection of known Hfq targets. mRNAs that are known RyhB targets might make a good control set, since growth conditions of cultures should be such that those mRNAs are not substantially repressed by RyhB. How many of these mRNAs are present at reasonable levels but have no signal in CLIP-seq? Or strong signal? From this analysis, combined with analysis of other known mRNA targets, is it possible to propose what factors modulate the efficiency of identification of Hfq binding sites by CLIP-seq? I think it's important to address this, especially since the authors are proposing to incorporate CLIP-seq data with computational analysis as a way to improve target predictions. It would be important to know rate of false negatives/false positives for CLIP-seq.
• Data represented in Fig. 4C should be shown in supplementary material, especially the list of targets predicted by CopraRNA that also have Hfq peaks. It is key to understand how the authors are defining "true positives." I assume the meaning is targets that have been defined by previous studies as being directly regulated by base pairing with the sRNA. However, the authors go on to show that some of the new targets (e.g., mglB) are, in fact, "true positives."
• Related to point above: How much can you extrapolate based on confirmation of one new sRNA-mRNA pair? Are the rest of the CopraRNA-predicted targets that have a corresponding Hfq peak now assumed to be true positives? If not, why not?
• For CsrA, I had the same question as above for Hfq, namely, Is there a way to quantify (even relative to other peaks) which are strong versus weak signals? Does this correlate with observed intensity of regulation (e.g., prg genes where some regulated and some are not?
• Could the regulation of SPI1 genes be explained by the known regulation of hilD? Why/why not? This could be addressed by expressing hilD from a construct that is not subject to CsrA-dependent regulation and monitoring CsrA effects on prg/sip genes in this background.
• Is there a control for Hfq-FLAG specificity/activity? Does the FLAG epitope alter function? Same for CsrA?

Referee #3:

This paper reports a transcriptome-wide analysis based on in vivo UV-crosslinking method with RNA deep sequencing (CLIP-seq) on the binding sites within RNAs of two RNA-binding proteins, Hfq and CsrA, in Salmonella. The study faithfully captured the known structural features of Hfq and CsrA binding sites within RNA molecules, by confirming the conclusion obtained by previous biochemical and structural studies on several sRNAs. Namely, Hfq preferentially binds RNA sequences corresponding to Rho-independent terminators of sRNAs and mRNAs while CsrA preferentially binds GGA sequences in loop regions in mRNAs and sRNAs. Thus, the CLIP-seq protocol described in this paper is quite useful to study RNA-protein interactions in living bacterial cells. In addition, the global analysis of Hfq-binding sites has convincingly established that Hfq preferentially binds 5' to sRNA target sites in mRNAs, and 3' to seed sequences in sRNAs, supporting a model how Hfq stimulates the sRNA-mRNA base pairing. The identification of new CsrA-binding sites in mRNAs indicates that CsrA is a regulator of Salmonella virulence genes. In

conclusion, the present study expanded our view regarding the function and mechanism of two well-studied RNA-binding proteins. The experiments are well designed, and executed thoroughly and carefully. The data and arguments are mostly clear and convincing to support the conclusion.

1st Revision - authors' response 26 January 2016

Referee #1:

In this manuscript Holmqvist and coworkers determine the RNA binding sites of two major bacterial post-transcriptional regulators, Hfq and CsrA, in Salmonella Typhimurium. Using CLIP-seq they identify RNA preferences and structural constraints of Hfq and CsrA. The binding site maps for Hfq interacting sRNAs and their target mRNAs support a model for Hfq as a mediator of RNA duplex formation, and provide new insight into improving sRNA target prediction.

The manuscript is well written and the results clearly presented. The conclusions are supported by experimental and computational results. The research topic and findings are of broader interest. Before considering this manuscript for publication, some issues should be addressed.

The authors should mention that they performed biological Hfq CLIP replicates. Since multiple replicates were generated the reproducibility and variance of these CLIP replicates should be assessed. This should not only been done at the transcript level, but also for binding sites. Based on the description in the methods sections reads from three replicates were pooled and the peak calling was performed. If this was the case, why not using more stringent criteria to identify peaks that have sequence coverage from at least 2 of the three replicates.

Reply: We thank reviewers 1 and 2 for pointing this out and agree on the need for a more detailed description of the statistical analysis of the CLIP-seq data for determination of binding sites (peaks). For clarity, the analysis was done as follows. The CLIP-seq experiments for Hfq and CsrA were performed in three biological replicates. For each bacterial culture (replicate), both crosslinked and non-crosslinked samples were prepared. Read clusters (and subsequent peaks) were defined from the union of reads from the three crosslinked samples. Each peak region was then tested for significant read enrichment in the crosslinked samples using DEseq2. In this second step, all libraries were considered individually (i.e. not pooled) and must show a reproducible read enrichment in crosslinked libraries. Peaks with an FDR-adjusted p-value (q-value) ≤ 0.1 were considered significant and were used for all downstream analysis.

To better explain the peak calling approach we have made several changes/additions in the manuscript. In Figure 1, a new panel has been added (1C) that graphically explains how the peak calling was performed. This panel is accompanied by additional text in Results. We have also re-written the peak calling part of the Materials & Methods section for clarity. In addition, Table EV1 (Hfq peaks) and Table EV4 (CsrA peaks) now include read counts, fold-changes and p-values for each peak. Correlations between sequencing libraries are provided in Appendix Figure S5. *In*

In Figure 2A the authors should also plot the peak density of peaks with containing T>C mutations.

Reply: The density of peaks with T>C mutations has been added to Figure 2B.

The authors argue that Hfq RNA binding data could increase the success of sRNA-target mRNA predictions. Using a mglB-gfp reporter they show that Spot42 represses a constitutively transcribed mglB-gfp fusion. To confirm that repression is mediated through an interaction of Spo42 with the mglB site ,identified by CLIP, the authors should mutate the reporter and show a direct effect on mglB-gfp expression.

Reply: We have added a panel in Figure 4 (panel H) showing an experiment in which the effect of disruptive and compensatory mutations in the predicted interaction between Spot42 and mglB-gfp was tested. The results experimentally validate the CopraRNA prediction that Spot42 regulates mglB through base-pairing: a GGG to CCC mutation in Spot42 nullifies repression of the *mglB::gfp*

fusion, an additional compensatory CCC to GGG mutation in the predicted target region in *mglB* restores it fully.

As for the analysis of the Hfq CLIP-seq data the reproducibility and variance of CsrA CLIP replicates should be assessed.

Reply: Please see our comment on the statistical analysis of CLIP-seq replicates above.

The authors show using a sopD2-gfp reporter that CsrA may directly repress the sopD2-gfp fusion. The reporter assays indicate that CsrA is involved in the regulation. To show that CsrA regulation is mediated by the interactions with the CLIP identified site in sopD2 the authors should mutate (or delete) the GGA motif in the binding site to examine the contribution of this motif.

Reply: Figure 7 now includes an experiment where the effect of mutating two GGA motifs in a CsrA peak in sopD2 was tested. The results show that the GGA sequences in question are required for the CsrA-dependent regulation of *sopD2*.

Lastly, the authors discuss the differences in RNA binding results from their study compared to Hfq CLIP data by Tree et al. at the molecular level. The question that arises is, whether targets are similar at the transcript level in E. coli and Salmonella Typhimurium.

Reply: We agree that an analysis of the conservation of Hfq binding sites would be very interesting. As described in the Discussion section, we think that differences between our protocol and the one in Tree et al led to rather distinct results. In addition, we note that their and our experiments were also conducted at different growth phases (we used early stationary phase to enable detection of RNAs important for invasion, whereas Tree et al. used cultures from exponential phase). We therefore believe that a comparative analysis between the data sets would be confounded by these technical and biological differences. That is, we would not be able to discriminate differences due to the protocols applied from differences in Hfq-binding patterns in the respective organisms.

Referee #2:

This manuscript describes the use of CLIP-seq in Salmonella to globally map protein-RNA interactions for two important RNA binding proteins, Hfq and CsrA. The data sets confirm interactions with known mRNA targets of these two proteins, and yields consensus binding site motifs that are consistent with published work. In addition, new putative targets are identified, and the authors do some experimental work to confirm that the binding sites found on some new targets are responsible for physiologically-relevant regulatory effects. The study provides a valuable new experimental model for genome-wide identification of protein-RNA interactions. In addition, the work suggests some general principles relevant to the activities of these two RNA binding proteins, e.g., the frequent location of Hfq binding sites 5' relative to sRNA binding sites on mRNA targets. Moreover, the results suggest that inclusion of Hfq binding site data could be used to improve computational prediction of sRNA-mRNA interactions. Overall, the study is extremely valuable, well written, and will be of broad interest. I am concerned about the qualitative nature of most of the observations, and think the study would be much improved if the authors could provide some context for or relative quantification of Hfq/CsrA interactions with the many different mRNA targets. My specific comments are below.

• The SPIs are more AT-rich, and this might increase non-specific interaction of Hfq with AU-rich RNAs derived from SPI1. Could this account for increased signal for Hfq across the SPIs?

Reply: We thank the referee for pointing this out and fully agree that the observed high density of Hfq binding sites in the SPIs may largely result from a higher AU content in these regions. This is also in line with our previous results with Hfq coIP (Sittka et al. 2008 PLoS Genet., Chao et al. 2012 EMBO J). A comment has been added in the Discussion on page 12.

• The Hfq consensus from peaks in 3' UTRs and sRNAs does resemble half of a GC-rich hairpin followed by a run of Us, as you would find in Rho-independent terminators. How frequently do the peaks containing this consensus contain the upstream half of the hairpin? The density of peaks around the terminator (Fig. 2H) does suggest that most of these motifs represent the real terminator

of the sRNA or mRNA, but this was not exactly clear.

Reply: To test for the presence of the upstream half of the stem, CMfinder analysis was carried out on the 3'UTR Hfq peaks. The resulting motif is presented in Figure EV1 and includes a full hairpin followed by a U-stretch, strengthening the conclusion that the motif shown in Figure 2H reflects the 3' part of Rho-independent terminators.

• My major concern is the lack of relative quantification of Hfq/CsrA interactions with different targets. The description of how peaks were called (in Materials and Methods) was complex and hard to follow. Some simplified description of what the authors mean by "significant peak" would be very helpful, as would a description of the range of peak size (read abundance?), normalized to total mRNA abundance, perhaps?

Reply: As stated above in the reply to reviewer 1, we fully agree that the description of the peak calling procedure required improvement. To make this part easier to follow, including how significance was determined, we have:

• added a panel to Figure 1 with a graphical description of the peak calling approach,

• changed the main text to clarify how the peak calling was done (page 6),

• amended the text describing the peak calling in the materials and methods,

• added read counts, fold changes and p-values for each peak to Table EV1 and EV4.

Regarding the normalization of the CLIP-seq data: As crosslinking may depend on many factors, e.g. distance between the molecules, lifetime of the complexes, the RNA sequence and the structural context, normalizing CLIP-seq data to RNA abundance may not necessarily inform on affinity.

• Please comment on the sensitivity of detection of known Hfq targets. mRNAs that are known RyhB targets might make a good control set, since growth conditions of cultures should be such that those mRNAs are not substantially repressed by RyhB. How many of these mRNAs are present at reasonable levels but have no signal in CLIP-seq? Or strong signal? From this analysis, combined with analysis of other known mRNA targets, is it possible to propose what factors modulate the efficiency of identification of Hfq binding sites by CLIP-seq? I think it's important to address this, especially since the authors are proposing to incorporate CLIP-seq data with computational analysis as a way to improve target predictions. It would be important to know rate of false negatives/false positives for CLIP-seq.

Reply: To address this, we performed an analysis of published total RNA-seq data for the same *Salmonella* strain (Kröger et al. 2013 Cell Host & Microbe). Briefly, we detected an Hfq peak in approximately 30% of all mRNAs previously reported to be regulated by sRNAs (Table S1 in Wright et al. 2013 PNAS) and with a TPM expression value >10 in the growth condition used for the CLIP-seq experiments here. This is remarkably sensitive given that first, many of the mRNAs may not be regulated by their cognate sRNA in the assayed condition (RyhB would be such an example); second, as discussed in the manuscript, our data may be biased towards U-rich Hfq binding sites, disfavoring the detection of A-rich binding events in mRNAs. Nevertheless, using the CLIP-seq data, the rate of true positives in CopraRNA predictions was strongly improved, suggesting that the CLIP-seq protocol yields data of high specificity, while sensitivity may be improved by the inclusion of more growth conditions. As computational predictions of sRNA:mRNA interactions generally suffer from an abundance of false positives (i.e. low specificity; Pain et al. 2015 RNA Biology), we believe our CLIP-seq data will indeed improve the prediction of mRNA targets. We have added a comment on page 14 to further clarify this issue.

• Data represented in Fig. 4C should be shown in supplementary material, especially the list of targets predicted by CopraRNA that also have Hfq peaks. It is key to understand how the authors are defining "true positives." I assume the meaning is targets that have been defined by previous studies as being directly regulated by base pairing with the sRNA. However, the authors go on to show that some of the new targets (e.g., mglB) are, in fact, "true positives."

Reply: Table EV3 lists the targets predicted by CopraRNA that also have Hfq peaks. For each previously validated target ("true positive") the original reference is given.

• Related to point above: How much can you extrapolate based on confirmation of one new sRNA-mRNA pair? Are the rest of the CopraRNA-predicted targets that have a corresponding Hfq peak now assumed to be true positives? If not, why not?

Reply: Detection of an Hfq binding site in an mRNA with a high-ranking CopraRNA prediction indicates a possibility for sRNA regulation. We show in Figure 4 that adding Hfq binding data in fact increases the fraction of true positive predictions considerably. To determine the fraction of true positives, all predictions given in TableEV3 would need to be tested experimentally, which we believe is beyond the scope of this study. However, mining published data sets showed that 30% of the predicted targets in Table EV3 are differentially expressed upon deletion of the hfq in *Salmonella* (Sittka et al. PLOS Genetics 2008, Ansong et al. PLOS One 2009) or after pulse-induction of the predicted cognate sRNA (Papenfort et al. Mol Microbiol. 2009), suggesting that they may be under sRNA control.

• For CsrA, I had the same question as above for Hfq, namely, Is there a way to quantify (even relative to other peaks) which are strong versus weak signals? Does this correlate with observed intensity of regulation (e.g., prg genes where some regulated and some are not?

Reply: To allow for the comparison of peaks we have added read counts, fold-changes and p-values for all significant CsrA peaks in Table EV4. Since the level of regulation for each transcript with a peak is not known, we can not calculate the correlation between peak strength and regulation. However, we note that in the case of the prg genes, the peak strength (fold-change) follows that of the observed regulation (prgJ>prgI>prgK).

• Could the regulation of SPI1 genes be explained by the known regulation of hilD? Why/why not? This could be addressed by expressing hilD from a construct that is not subject to CsrA-dependent regulation and monitoring CsrA effects on prg/sip genes in this background.

Reply: The regulation of SPI1 genes was tested using translational gfp-fusions driven by a heterologous PLtetO-1 promoter, which is not regulated in a CsrA-dependent manner (Figure EV4, lacZ-gfp fusion). Therefore, the reported regulation occurs on the post-transcriptional level, independent of transcriptional regulation by HilD.

• Is there a control for Hfq-FLAG specificity/activity? Does the FLAG epitope alter function? Same for CsrA?

Reply: We have added experiments addressing the possible effects of the FLAG epitope on bacterial growth and Hfq- or CsrA-dependent regulation. The tagged strains were found to behave identically to the wild type strain in these assays. The data is presented in Appendix Figure S4.

Referee #3:

This paper reports a transcriptome-wide analysis based on in vivo UV-crosslinking method with RNA deep sequencing (CLIP-seq) on the binding sites within RNAs of two RNA-binding proteins, Hfq and CsrA, in Salmonella. The study faithfully captured the known structural features of Hfq and CsrA binding sites within RNA molecules, by confirming the conclusion obtained by previous biochemical and structural studies on several sRNAs. Namely, Hfq preferentially binds RNA sequences corresponding to Rho-independent terminators of sRNAs and mRNAs while CsrA preferentially binds GGA sequences in loop regions in mRNAs and sRNAs. Thus, the CLIP-seq protocol described in this paper is quite useful to study RNA-protein interactions in living bacterial cells. In addition, the global analysis of Hfq-binding sites has convincingly established that Hfq preferentially binds 5′ to sRNA target sites in mRNAs, and 3′ to seed sequences in sRNAs, supporting a model how Hfq stimulates the sRNA-mRNA base pairing. The identification of new CsrA-binding sites in mRNAs indicates that CsrA is a regulator of Salmonella virulence genes. In conclusion, the present study expanded our view regarding the function and mechanism of two well-studied RNA-binding proteins. The experiments are well designed, and executed thoroughly and carefully. The data and arguments are mostly clear and convincing to support the conclusion.

Reply: We take this to mean that this referee recommends publication of our manuscript without further changes.

2nd Editorial Decision 16 February 2016

Thank you for submitting a revised version of your manuscript to The EMBO Journal. It has now been seen by two of the original referees whose comments are shown below.

As you will see they both find that all criticisms have been sufficiently addressed and they recommend the manuscript for publication. However, before we can officially accept the study there are a few remaining editorial issues that I would ask you to address in a final revision:

-> Please include a brief statement on author contributions and conflict of interest in the manuscript file

-> Please fill out and include an author checklist as listed in our online guidelines (http://emboj.embopress.org/authorguide)

-> Please include a brief legend/description for each of the Expanded View tables. These can simply be inserted as a separate tab in each sheet. version)

-> We generally encourage the publication of source data, particularly for electrophoretic gels and blots, with the aim of making primary data more accessible and transparent to the reader. We would need 1 file per figure (which can be a composite of source data from several panels) in jpg, gif or PDF format, uploaded as "Source data files". The gels should be labelled with the appropriate figure/panel number, and should have molecular weight markers; further annotation would clearly be useful but is not essential. These files will be published online with the article as a supplementary "Source Data".

-> Papers published in The EMBO Journal include a 'Synopsis' to further enhance discoverability. Synopses are displayed on the html version of the paper and are freely accessible to all readers. The synopsis includes a short standfirst - written by the handling editor - as well as 2-5 one-sentence bullet points that summarise the paper and are provided by the authors. I would therefore ask you to include your suggestions for bullet points.

-> In addition, I would encourage you to provide an image for the synopsis. This image should provide a rapid overview of the question addressed in the study but still needs to be kept fairly modest since the image size cannot exceed 550x400 pixels.

Thank you again for giving us the chance to consider your manuscript for The EMBO Journal, I look forward to receiving your final revision.

Referee #1:

The authors addressed all issues. The manuscript should be considered for publication in EMBO J.

Referee #2:

The authors addressed my comments on the previous submission very well. This is an outstanding piece of work, and will be well received by the community.

EMBO PRESS

YOU MUST COMPLETE ALL CELLS WITH A PINK BACKGROUND ↓

PLEASE NOTE THAT THIS CHECKLIST WILL BE PUBLISHED ALONGSIDE YOUR PAPER

Corresponding Author Name: Jörg Vogel, Rolf Backofen
Journal Submitted to: EMBO Journal
Manuscript Number: EMBOJ-2015-93360

Reporting Checklist For Life Sciences Articles (Rev. July 2015)

This checklist is used to ensure good reporting standards and to improve the reproducibility of published results. These guidelines are consistent with the Principles and Guidelines for Reporting Preclinical Research issued by the NIH in 2014. Please follow the journal's authorship guidelines in preparing your manuscript.

A- Figures

1. Data

The data shown in figures should satisfy the following conditions:

- the data were obtained and processed according to the field's best practice and are presented to reflect the results of the experiments in an accurate and unbiased manner.
- figure panels include only data points, measurements or observations that can be compared to each other in a scientifically meaningful way.
- graphs include clearly labeled error bars for independent experiments and sample sizes. Unless justified, error bars should not be shown for technical replicates.
- if n< 5, the individual data points from each experiment should be plotted and any statistical test employed should be justified
- Source Data should be included to report the data underlying graphs. Please follow the guidelines set out in the author ship guidelines on Data Presentation.

2. Captions

Each figure caption should contain the following information, for each panel where they are relevant:

- a specification of the experimental system investigated (eg cell line, species name).
- the assay(s) and method(s) used to carry out the reported observations and measurements
- an explicit mention of the biological and chemical entity(ies) that are being measured.
- an explicit mention of the biological and chemical entity(ies) that are altered/varied/perturbed in a controlled manner.

- the exact sample size (n) for each experimental group/condition, given as a number, not a range;
- a description of the sample collection allowing the reader to understand whether the samples represent technical or biological replicates (including how many animals, litters, cultures, etc.).
- a statement of how many times the experiment shown was independently replicated in the laboratory.
- definitions of statistical methods and measures:
  - common tests, such as t-test (please specify whether paired vs. unpaired), simple $\chi 2$ tests, Wilcoxon and Mann-Whitney tests, can be unambiguously identified by name only, but more complex techniques should be described in the methods section;
  - are tests one-sided or two-sided?
  - are there adjustments for multiple comparisons?
  - exact statistical test results, e.g., P values = x but not P values < x;
  - definition of 'center values' as median or average;
  - definition of error bars as s.d. or s.e.m.

Any descriptions too long for the figure legend should be included in the methods section and/or with the source data.

Please ensure that the answers to the following questions are reported in the manuscript itself. We encourage you to include a specific subsection in the methods section for statistics, reagents, animal models and human subjects.

In the pink boxes below, provide the page number(s) of the manuscript draft or figure legend(s) where the information can be located. Every question should be answered. If the question is not relevant to your research, please write NA (non applicable).

B- Statistics and general methods

Please fill out these boxes ↓ (Do not worry if you cannot see all your text once you press return)

1.a. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size?	NA
1.b. For animal studies, include a statement about sample size estimate even if no statistical methods were used.	NA
2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established?	NA
3. Were any steps taken to minimize the effects of subjective bias when allocating animals/samples to treatment (e.g. randomization procedure)? If yes, please describe.	NA
For animal studies, include a statement about randomization even if no randomization was used.	NA
4.a. Were any steps taken to minimize the effects of subjective bias during group allocation or/and when assessing results (e.g. blinding of the investigator)? If yes please describe.	NA
4.b. For animal studies, include a statement about blinding even if no blinding was done	NA
5. For every figure, are statistical tests justified as appropriate?	NA
Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it.	NA
Is there an estimate of variation within each group of data?	NA
Is the variance similar between the groups that are being statistically compared?	NA

C- Reagents

USEFUL LINKS FOR COMPLETING THIS FORM

http://www.antibodypedia.com
http://1degreebio.org
http://www.equator-network.org/reporting-guidelines/improving-bioscience-research-repo

http://grants.nih.gov/grants/olaw/olaw.htm
http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/Useofanimals/index.htm
http://ClinicalTrials.gov
http://www.consort-statement.org
http://www.consort-statement.org/checklists/view/32-consort/66-title

http://www.equator-network.org/reporting-guidelines/reporting-recommendations-for-tun

http://datadryad.org

http://figshare.com

http://www.ncbi.nlm.nih.gov/gap

http://www.ebi.ac.uk/ega

http://biomodels.net/

http://biomodels.net/miriam/
http://jjj.biochem.sun.ac.za
http://oba.od.nih.gov/biosecurity/biosecurity_documents.html
http://www.selectagents.gov/

6. To show that antibodies were profiled for use in the system under study (assay and species), provide a citation, catalog number and/or clone number, supplementary information or reference to an antibody validation profile. e.g., Antibodypedia (see link list at top right), 1DegreeBio (see link list at top right).	Catalog numbers for antibodies are given in the Material and Methods section of the manuscript
7. Identify the source of cell lines and report if they were recently authenticated (e.g., by STR profiling) and tested for mycoplasma contamination.	NA

* for all hyperlinks, please see the table at the top right of the document

D- Animal Models

8. Report species, strain, gender, age of animals and genetic modification status where applicable. Please detail housing and husbandry conditions and the source of animals.	NA
9. For experiments involving live vertebrates, include a statement of compliance with ethical regulations and identify the committee(s) approving the experiments.	NA
10. We recommend consulting the ARRIVE guidelines (see link list at top right) (PLoS Biol. 8(6), e1000412, 2010) to ensure that other relevant aspects of animal studies are adequately reported. See author guidelines, under 'Reporting Guidelines'. See also: NIH (see link list at top right) and MRC (see link list at top right) recommendations. Please confirm compliance.	NA

E- Human Subjects

11. Identify the committee(s) approving the study protocol.	NA
12. Include a statement confirming that informed consent was obtained from all subjects and that the experiments conformed to the principles set out in the WMA Declaration of Helsinki and the Department of Health and Human Services Belmont Report.	NA
13. For publication of patient photos, include a statement confirming that consent to publish was obtained.	NA
14. Report any restrictions on the availability (and/or on the use) of human data or samples.	NA
15. Report the clinical trial registration number (at ClinicalTrials.gov or equivalent), where applicable.	NA
16. For phase II and III randomized controlled trials, please refer to the CONSORT flow diagram (see link list at top right) and submit the CONSORT checklist (see link list at top right) with your submission. See author guidelines, under 'Reporting Guidelines'. Please confirm you have submitted this list.	NA
17. For tumor marker prognostic studies, we recommend that you follow the REMARK reporting guidelines (see link list at top right). See author guidelines, under 'Reporting Guidelines'. Please confirm you have followed these guidelines.	NA

F- Data Accessibility

18. Provide accession codes for deposited data. See author guidelines, under 'Data Deposition'. Data deposition in a public repository is mandatory for: a. Protein, DNA and RNA sequences b. Macromolecular structures c. Crystallographic data for small molecules d. Functional genomics data e. Proteomics and molecular interactions	Sequencing data is available via Gene Expression Omnibus under accession number GSE74425
19. Deposition is strongly recommended for any datasets that are central and integral to the study; please consider the journal's data policy. If no structured public repository exists for a given data type, we encourage the provision of datasets in the manuscript as a Supplementary Document (see author guidelines under 'Expanded View' or in unstructured repositories such as Dryad (see link list at top right) or Figshare (see link list at top right).	NA
20. Access to human clinical and genomic datasets should be provided with as few restrictions as possible while respecting ethical obligations to the patients and relevant medical and legal issues. If practically possible and compatible with the individual consent agreement used in the study, such data should be deposited in one of the major public access-controlled repositories such as dbGAP (see link list at top right) or EGA (see link list at top right).	NA
21. As far as possible, primary and referenced data should be formally cited in a Data Availability section. Please state whether you have included this section. Examples: **Primary Data** Wetmore KM, Deutschbauer AM, Price MN, Arkin AP (2012). Comparison of gene expression and mutant fitness in Shewanella oneidensis MR-1. Gene Expression Omnibus GSE39462 **Referenced Data** Huang J, Brown AF, Lei M (2012). Crystal structure of the TRBD domain of TERT and the CR4/5 of TR. Protein Data Bank 4O26 AP-MS analysis of human histone deacetylase interactions in CEM-T cells (2013). PRIDE PXD000208	NA
22. Computational models that are central and integral to a study should be shared without restrictions and provided in a machine-readable form. The relevant accession numbers or links should be provided. When possible, standardized format (SBML, CellML) should be used instead of scripts (e.g. MATLAB). Authors are strongly encouraged to follow the MIRIAM guidelines (see link list at top right) and deposit their model in a public database such as Biomodels (see link list at top right) or JWS Online (see link list at top right). If computer source code is provided with the paper, it should be deposited in a public repository or included in supplementary information.	NA

G- Dual use research of concern

23. Could your study fall under dual use research restrictions? Please check biosecurity documents (see link list at top right) and list of select agents and toxins (APHIS/CDC) (see link list at top right). According to our biosecurity guidelines, provide a statement only if it could.	No